

RESEARCH

Open Access



Differentiating nontuberculous mycobacterium pulmonary disease from pulmonary tuberculosis through the analysis of the cavity features in CT images using radiomics

Qinghu Yan¹, Wuzhang Wang¹, Wenlong Zhao¹, Liping Zuo², Dongdong Wang², Xiangfei Chai³ and Jia Cui^{1*}

Abstract

Objective: To differentiate nontuberculous mycobacteria (NTM) pulmonary diseases from pulmonary tuberculosis (PTB) by analyzing the CT radiomics features of their cavity.

Methods: 73 patients of NTM pulmonary diseases and 69 patients of PTB with the cavity in Shandong Province Chest Hospital and Qilu Hospital of Shandong University were retrospectively analyzed. 20 patients of NTM pulmonary diseases and 20 patients of PTB with the cavity in Jinan Infectious Disease Hospital were collected for external validation of the model. 379 cavities as the region of interest (ROI) from chest CT images were performed by 2 experienced radiologists. 80% of cavities were allocated to the training set and 20% to the validation set using a random number generated by a computer. 1409 radiomics features extracted from the Huiying Radcloud platform were used to analyze the two kinds of diseases' CT cavity characteristics. Feature selection was performed using analysis of variance (ANOVA) and least absolute shrinkage and selection operator (LASSO) methods, and six supervised learning classifiers (KNN, SVM, XGBoost, RF, LR, and DT models) were used to analyze the features.

Results: 29 optimal features were selected by the variance threshold method, K best method, and Lasso algorithm, and the ROC curve values are obtained. In the training set, the AUC values of the six models were all greater than 0.97, 95% CI were 0.95–1.00, the sensitivity was greater than 0.92, and the specificity was greater than 0.92. In the validation set, the AUC values of the six models were all greater than 0.84, 95% CI were 0.76–1.00, the sensitivity was greater than 0.79, and the specificity was greater than 0.79. In the external validation set, The AUC values of the six models were all greater than 0.84, LR classifier has the highest precision, recall and F1-score, which were 0.92, 0.94, 0.93.

Conclusion: The radiomics features extracted from cavity on CT images can provide effective proof in distinguishing the NTM pulmonary disease from PTB, and the radiomics analysis shows a more accurate diagnosis than the radiologists. Among the six classifiers, LR classifier has the best performance in identifying two diseases.

Keywords: Radiomics, Cavity, CT, Nontuberculous mycobacteria, Pulmonary tuberculosis

Background

Globally, the incidence of nontuberculous mycobacteria (NTM) has been increasing rapidly in recent years [1, 2]. NTM, usually involving the lung, exhibits characteristics

*Correspondence: 584872500@qq.com

¹ Department of Radiology, Shandong Public Health Clinical Center, Jinan 250013, China

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

such as slow growth, long course, and high resistance to first-line anti-tuberculosis drugs, resulting in a low total cure rate [3]. To improve the treatment effect, pretreatment diagnosis is very crucial. However, the CT features and the clinical manifestations of the NTM pulmonary disease are similar to those of pulmonary tuberculosis (PTB), so it is very difficult to identify these two diseases simply via CT or clinical features. At present, the main methods of distinguishing them include the culture of mycobacterium in sputum and the identification of the species. However, these methods are time- and resource-consuming and require higher laboratory standards, which cannot be provided in many local hospitals [4]. Moreover, even with a positive culture result, there is still a need for further identification of the species. Once being suspected of PTB with positive sputum AFB, a patient will be administered an empirical anti-TB medicine for treatment when the test is ongoing to identify the bacteria. This means that some patients may receive unnecessary treatment, which may consequently result in the risk of adverse drug reactions, with its accompanying unnecessary medical expenses [5]. Therefore, the early diagnosis of NTM pulmonary disease can improve the quality of patients' life, and contribute to disease treatment. Some studies [6–9] have shown that there are usually changes in the imaging of NTM pulmonary disease. A Tomography (CT) scan can be used to identify the changes in the lung areas including cavity formation, parenchymal lesion, bronchiectasis, and tree-in-bud patterns [9, 10, 20], particularly the cavity, and the cavity is also a common imaging feature of chronic PTB [11, 12]. Unfortunately, the CT cavity images do not provide effective identification points in these two diseases. In recent years, radiomics, mining data features through high-throughput extraction have shown great potential in the diagnosis and differentiation of lung diseases, such as pulmonary nodules and early lung cancer [13, 14], and this may provide a feasible method of identifying the NTM pulmonary disease from PTB. In addition, the pathological difference in the cavity between the two diseases has been proven [15], which increases their possibility of identification. Thus, this study investigates the value of CT radiomics in the cavity analysis and identification of the NTM and PTB pulmonary diseases. To explore the significance of this technique in the differential diagnosis of the two diseases, select excellent classifiers, and provide a new and simple diagnostic method for clinical treatment.

Materials and methods

Patient information

The study retrospectively analyzed 89 patients with NTM pulmonary disease and other 100 cases of patients

with PTB [16] in CT imaging in the Shandong Provincial Chest Hospital and Qilu Hospital of Shandong University from February 2013 to March 2018. In addition, 27 patients with NTM pulmonary disease and other 30 cases of patients with PTB [16] in CT imaging at The Jinan Infectious Disease Hospital were collected for external validation of the model. The content of this study was approved by the ethics committee of Shandong Provincial Chest Hospital (Additional files 1, 2, 3). All patients with lung changes underwent a plain chest scan using Philips 64-slice spiral CT scanner before treatment, and the quality correction was met before the CT scan. The NTM pulmonary disease or PTB patient was laid in a supine position, and CT scanned from the tip of the lungs to the level of the thoracic cavity with the patient holding his breath. The CT scanning parameters were as follows: the diameter of the inspected detector, 64×0.625 mm; the rotation time, 0.5 s; the pitch, 1.375; the tube voltage, 120 kV; the tube current, 250–400 mA, which was modulated using an automatic tube current; and the thickness of the slice, 5 mm.

The inclusion criteria: (1) All cases of patients with NTM pulmonary disease who were diagnosed with sputum cultured twice and the strains were identified as the same pathogenic bacteria [12, 17]. All cases of patients with PTB who were identified by the new version of the tuberculosis diagnosis and tuberculosis classification standard issued by the China National Health Service Commission on November 9, 2017. The sputum smear of the same patient was positive for acid-fast bacilli at least once and was identified as *Mycobacterium tuberculosis* complex using the colloidal gold method. Also, clinical symptoms, laboratory examination, or imaging features suggest tuberculosis infection (2). The CT images (lung window) indicated the existence of a more than 3 mm cavity. (3) Corresponding anti-NTM or -PTB treatment was effective. (4) All images satisfied the analytical criteria for radiomics. Exclusion criteria: (1) the presence of lesions without cavities as depicted by the CT image (2) Some factors might influence the accuracy of the analysis including the CT images with motion artifacts, poor image quality, massive differences in scanning parameters, and inconsistent slice thicknesses. (3) The bullae and other cavities without a wall. (4) cases in which corresponding treatments had no effect. (5) The clinical data showed that the patient had the possibility of mixed infection. (6) The outer wall is difficult to be delineated because it was covered by the surrounding opacities or consolidation. After screening, 23 patients with NTM pulmonary disease and 41 patients with PTB were excluded, thus leaving a total of 73 patients with NTM pulmonary disease, including 47 males and 26 females, with an average age of 57.3 ± 14.0 years (ranged

from 23–86 years) and 69 patients with PTB, including 48 males and 21 females; mean age 45.7 ± 17.9 years (ranged from 17–90 years) to participate in this study. The external validation model had 20 patients with NTM pulmonary diseases (15 men and 5 women; mean age, 52.2 ± 16.2 years; range, 17–72 years), and 20 patients with PTB (12 men and 8 women; mean age, 43.8 ± 17.7 years; range, 19–76 years), with the cavity.

Characterization of lesions

All images and clinical data were uploaded to Radiance's Radcloud platform (Huiying Medical Technology Co., Ltd) for further study. Two radiologists (reader 1: Q.Y, six years of experience, and reader 2: S.Q, ten years of experience) with expertise in chest disease were invited to interpret and delineate the cavities in the lung window CT images (W 1500, L -500). The two readers were both blinded to the clinical and pathogenic information. The first and last slices were excluded to eliminate the errors due to the volume effect, then reader 1 manually sketched the ROI by delineating the outline of the entire cavity on all contiguous slices. The contour line was drawn on the outer edge of the cavity and should completely include the cavity and its wall. If there were ground-glass opacities or consolidation partly next to the cavity, the outline was drawn along the existing outer wall based on its thickness and the shape of the inner wall. If there were multiple cavities in the patient's lungs, each cavity was individually outlined as an independent ROI, and if two or more lung cavities were connected or multilocular, they would be treated as one cavity and drawn at the outer wall, also cavities with diameters less than 0.5 cm were not drawn because of possible delineating error. A total of 379 ROIs were manually mapped out

from the CT images of 182 patients. 142 patients were used for model training and internal validation, and 300 ROIs (164 NTM, 136 PTB) were labeled. 40 patients for external validation were labeled with 79 ROIs (49 NTM, 30 PTB). Subsequently, the second reader with a senior professional title then reviewed all the delineations. If the difference between the two readers was greater than or equal to 5%, the senior reader determined the cavity ROI boundary [18]. The contour lines of the cavity from the lung-window CT images are shown in Fig. 1a (NTM pulmonary disease) and Fig. 1b (PTB).

Features extraction

A total of 1409 quantitative imaging features were extracted from ROIs with the RadCloud platform. These features were divided into three groups: Group 1 (intensity statistics features) consisted of 270 features that quantitatively delineated the distribution of voxel intensities within the CT images through commonly used and basic metrics. Group 2 (shape- and size-based features) contained 16 three-dimensional features that reflected the shape and size of the region. Group 3 (texture features) calculated by grey level run-length matrices (GLRLM), grey level co-occurrence texture matrices (GLCM), gray level size zone matrix (GLSZM), neighbouring gray tone difference matrix (NGTDM), and gray level dependence matrix (GLDM), and 1123 textural features that could quantify region heterogeneity differences were classified into group 3 (texture features).

Features selection

As described above, high-throughput image features may be computed. However, not all the features were useful for a particular task. Therefore, in order to obtain

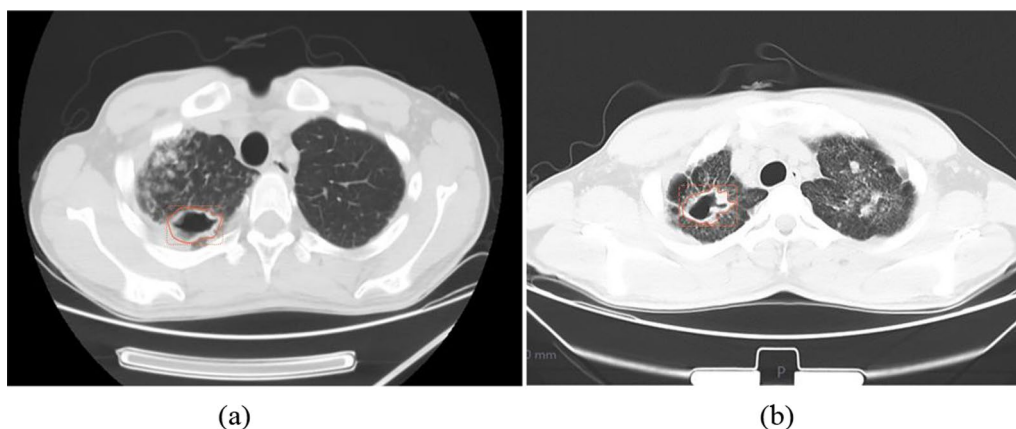


Fig. 1 **a** A patient (male, 54 years old) with a cavity of NTM pulmonary disease in the right superior lobe. Outline along the outer edge of the cavity and include the cavity wall. **b** A patient (male, 65 years old) with a similar cavity of PTB in the right superior lobe. Outline along the outer edge of the cavity and include the cavity wall.

the most relevant features, feature selection for the specific task was a necessary steps. Analysis of variance (ANOVA) and least absolute shrinkage and selection operator (LASSO) were used to remove redundant feature. ANOVA belongs univariate feature selection method, using P value < 0.05 as the significant differences in features between the two lesions. For the LASSO method, L1 regularizer was used as the cost function, the error value of cross validation was 5, and the maximum number of iterations was 1000.

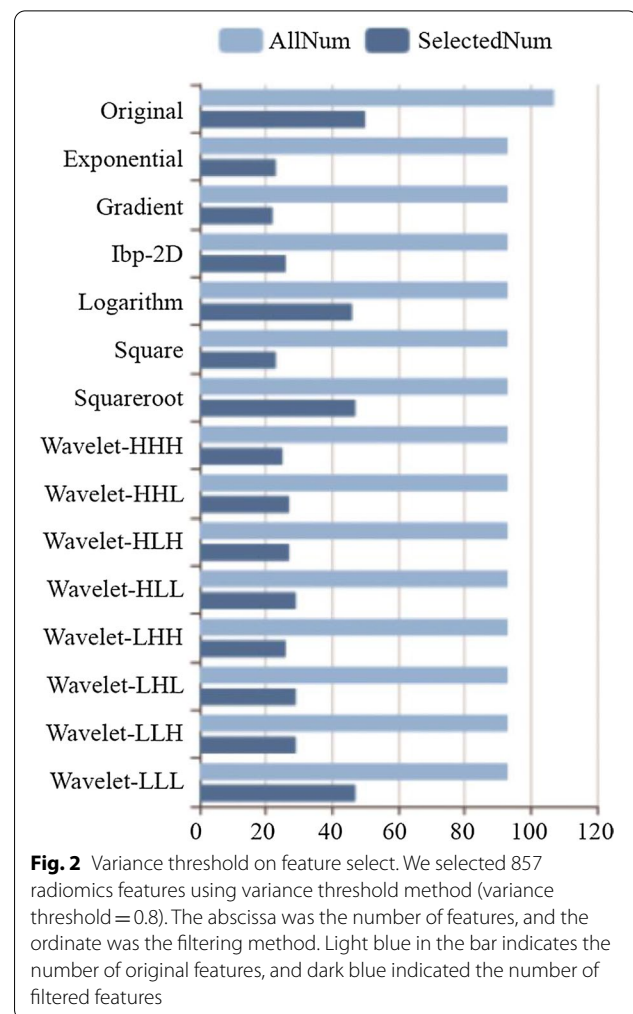
Statistical analysis

The statistical analysis was performed in the Radcloud platform. After feature selection, a total of 66 features identified were significantly correlated to differentiating NTM pulmonary disease from PTB. Based on the selected features, there were several supervised learning classifiers available for classification analysis. In this study, the radiomics-based models were constructed with 6 classifiers, k-NearestNeighbor (KNN), Support Vector Machin (SVM), eXtreme Gradient Boosting (XGBoost), Random Forest (RF), Logistic Regression (LR) and Decision tree(DT), and the validation set was used to evaluate the effectiveness of these models. To assess the predictive performance, the receiver operating characteristic (ROC) curve and the area under curve (AUC) were used both in training set and validation set respectively. And four indicators including P (precision = true positives / (true positives + false positives)), R (recall = true positives / (true positives + false negatives)), f1-score (f1-score = $P \cdot R \cdot 2 / (P + R)$), support (total number in test set) to evaluate the performance of classifier in this study.

Results

Firstly, 476 features were selected from 1409 features using the variance threshold method (Fig. 2). Then 349 features were selected with the ANOVA method (Fig. 3). Finally, we selected 29 optimal features with LASSO algorithm (Fig. 4). When training with KNN SVM, XGBoost, RF, LR and DT classifier, the AUC of training set and validation set were shown in Figs. 5 and 6.

Six classifiers were used to analyze the characteristics of the radiomic AUC, the 95% CI, sensitivity, and specificity of the training and verification sets are shown in Tables 1 and 2. The AUC values of the six models in the training set were all greater than 0.98, with a 95% CI ranging between 0.95 to 1.00, the sensitivity was greater than 0.92, with a specificity value greater than 0.92. In the validation set, the AUC values of the six models were all greater than 0.85, with a 95% CI of 0.76–1.00, the sensitivity was greater than 0.79, and the specificity was greater than 0.79 furthermore, the results showed that irrespective of either the training set or the truth



verification set, the AUC value was significantly high and greater than 0.85 in both instances. The AUC of the DT classifier was the lowest with a 95% CI of 0.76–0.94, sensitivity and specificity were greater than 0.79.

In the training set, the precision of the six models was greater than 0.90, the recall rate was greater than 0.92, the F1 score was greater than 0.93, and the support was 108 and 131 (Table 3). In the validation set, the precision of the six models was greater than 0.83 with a recall rate greater than 0.79, the F1 score was greater than 0.83, and the support was 28 and 33 (Table 4).

Six classifiers were same used to analyze the characteristics of radiomic AUC, 95% CI, sensitivity and specificity of external validation set were shown in Table 5. The AUC values of the six models were all greater than 0.84, 95% CI were 0.64–1.00, the sensitivity was greater than 0.61, and the specificity was greater than 0.61. It was found from the table data that external validation set, The AUC value was similar to the results of internal validation, but the sensitivity, specificity and 95% CI were lower

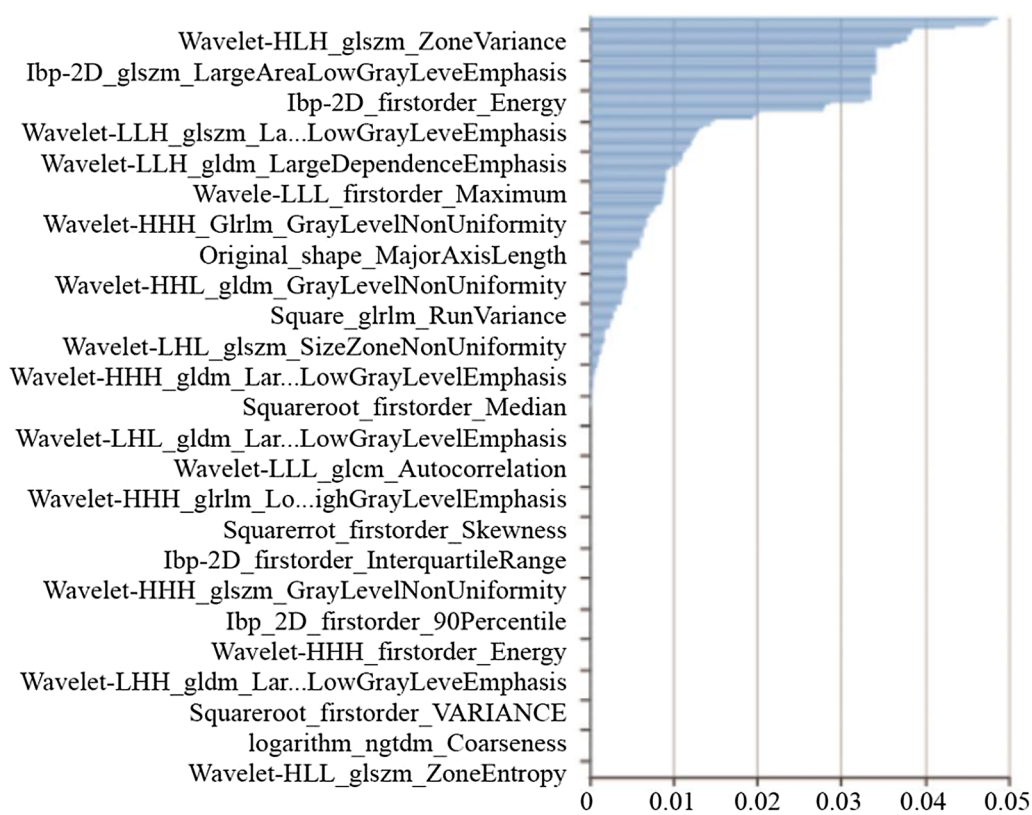


Fig. 3 ANOVA on feature selecting. We used ANOVA method to further select radiomics features, and 631 features were selected. The abscissa was the P -value of the feature

than those of internal validation. Among the six classifiers, LR classifier has the highest sensitivity, specificity and 95% CI, which were 0.94, 0.87, 0.84–0.96.

In the external validation set, the precision of the six models was greater than 0.57, the recall rate was greater than 0.61, the F1 score was greater than 0.68, and the support was 30 and 49 (Table 6). In the six classifiers, the LR classifier has the highest precision, recall, and F1-score.

Discussion

NTM pulmonary disease refers to lung diseases caused by mycobacteria except for *Mycobacterium tuberculosis* complex and *Mycobacterium leprae*. Generally, these bacteria exist mainly in water and soil mediums, with over 190 NTM types identified in previous studies [17, 19–21], Among which only a small number of bacteria are disease-causing agents. The incidence and prevalence of NTM pulmonary disease are increasing in some countries and regions [22–25], with some regions reporting higher cases of NTM than PTB [3, 19]. The identification of NTM and PTB pulmonary diseases requires complex pathogenic testing. And the simple and non-invasive chest CT imaging has a very important role in diagnosis

and differential diagnosis. The CT of NTM pulmonary disease shows a variety of signs such as cavity, and bronchiectasis, which makes it highly indistinguishable from PTB [26–28]. Studies have shown that cavity is a more common imaging manifestation in NTM pulmonary disease [29, 30]. There are plenty of literature results on the comparison of cavity types in pulmonary disease, with no consensus [8–11, 31]. Some studies [32] had shown that the cavities in NTM pulmonary disease are closer to the pleura of the lungs than those of PTB. Other studies [8, 9] pointed out that NTM pulmonary disease has thinner-walled cavities than PTB, with some researchers also reporting no statistically significant differences in the cavity wall results for the two diseases [6, 31]. This shows the existence of significant discrepancies in the use of conventional CT to identify the cavity features, especially its practical value.

Recent years have seen rapid developments in radiomics which converts medical images into high-dimensional images, and obtains quantifiable data features through quantitative high-throughput analysis, and then performs data analysis to provide clinical support. Radiomics has been shown to play a very important role in the diagnosis

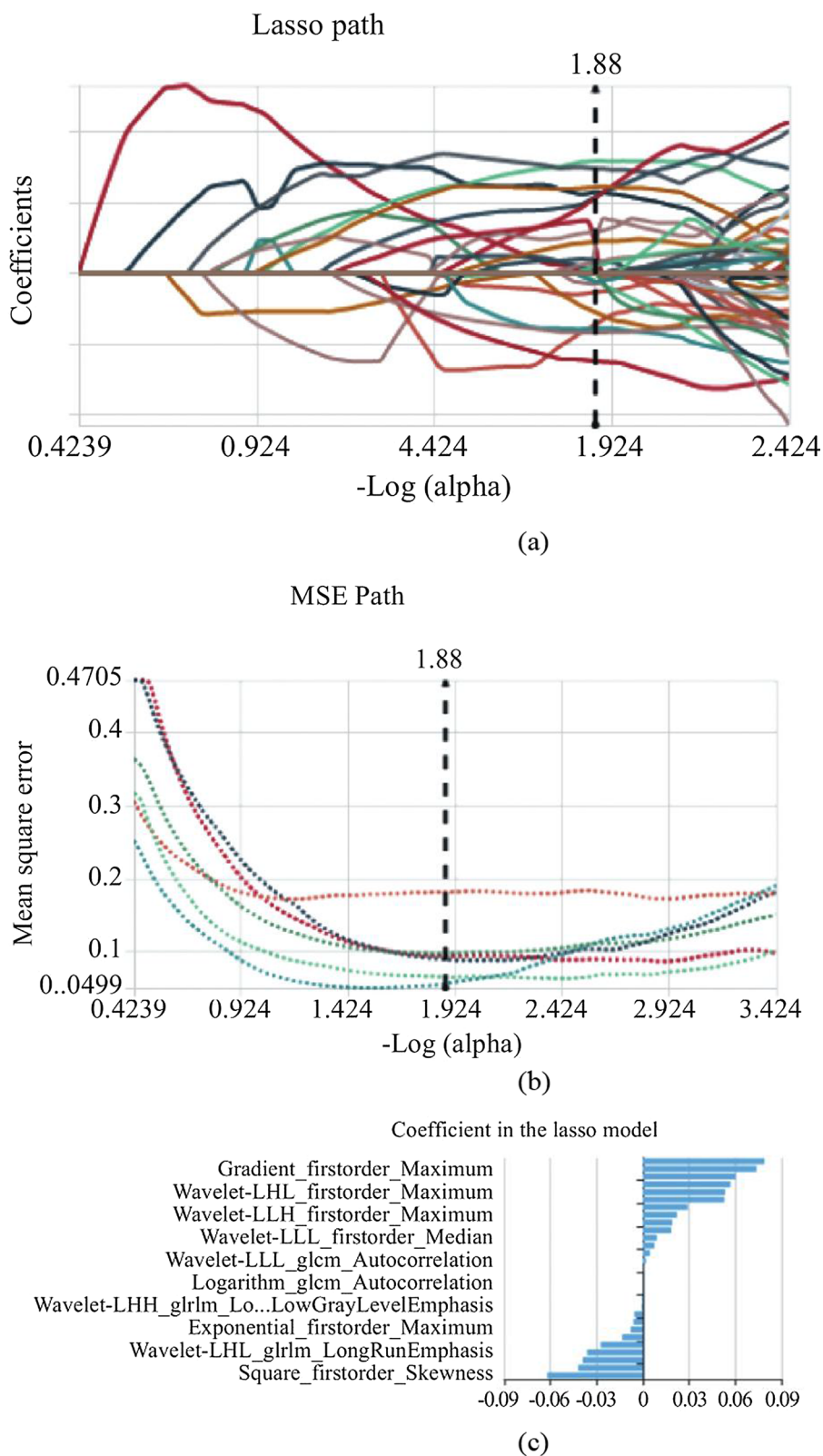


Fig. 4 Lasso algorithm on feature selecting. **a** Lasso path; **b** MSE path; **c** coefficients in Lasso model. Using Lasso model, 66 features which were correspond to the optimal alpha value were selected

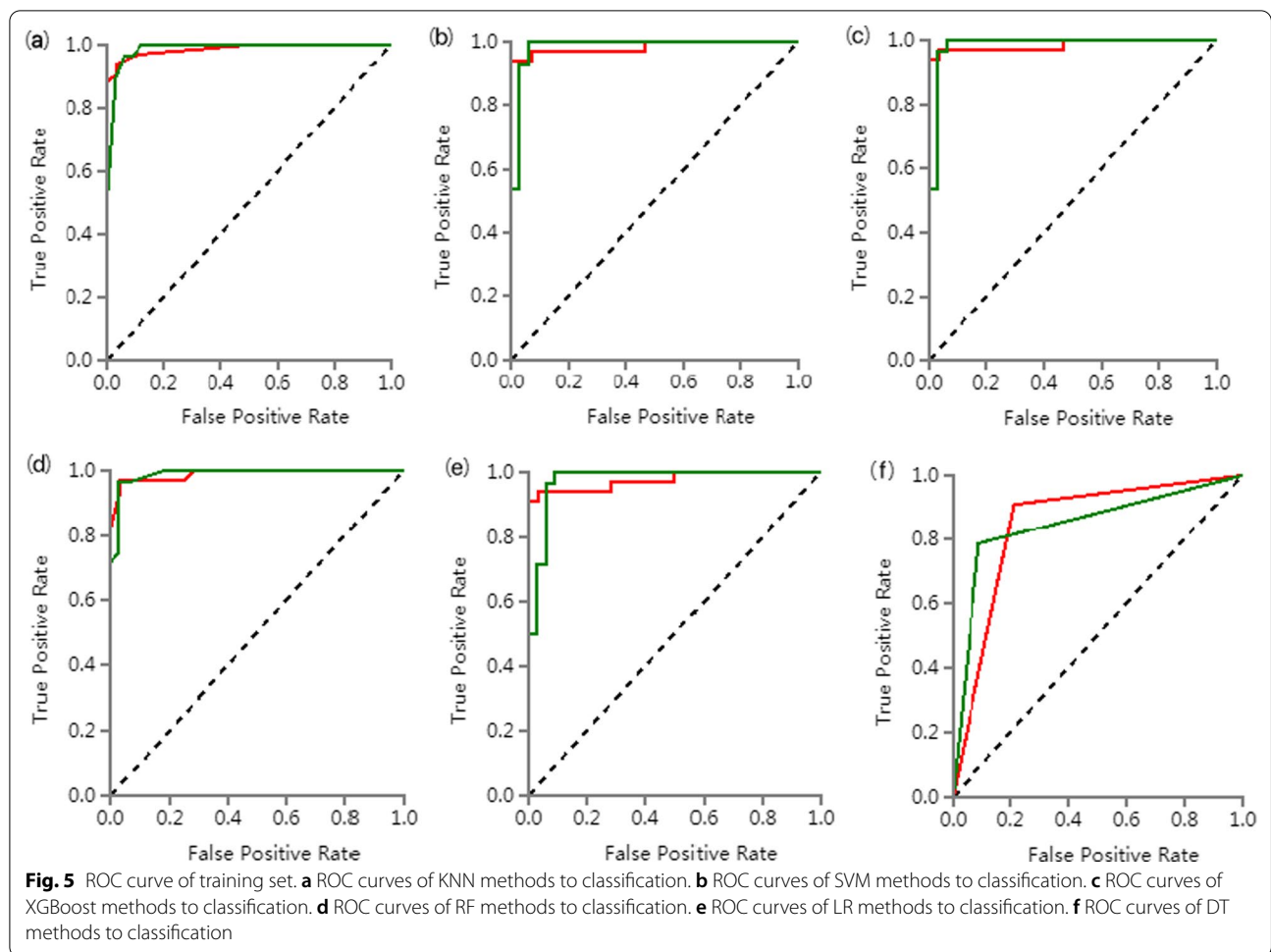


Table 1 ROC results with KNN, SVM, XGBoost, RF, LR and DT classifiers of training set

Classifiers	Category	AUC	95% CI	Sensitivity	Specificity
KNN	NTM	0.99	0.96–1.00	0.92	0.96
	TB	0.99	0.96–1.00	0.96	0.92
SVM	NTM	0.98	0.95–1.00	0.95	0.96
	TB	0.98	0.95–1.00	0.96	0.95
XGBoost	NTM	1.00	0.99–1.00	0.98	0.99
	TB	1.00	0.99–1.00	0.99	0.98
RF	NTM	1.00	0.99–1.00	0.99	1.00
	TB	1.00	0.99–1.00	1.00	0.99
LR	NTM	0.99	0.96–1.00	0.95	0.95
	TB	0.99	0.96–1.00	0.95	0.95
DT	NTM	1.00	1.00–1.00	1.00	1.00
	TB	1.00	1.00–1.00	1.00	1.00

t

and differential diagnosis of clinical pulmonary diseases, such as the screening of small pulmonary nodules, diagnosis of lung cancer, the guidance of the treatment of

pulmonary disease, and prognosis prediction [32–34]. Also, it has been well applied in the diagnosis of some common infectious diseases [35, 36]. However, there are relatively few studies on diffuse pulmonary disease [35, 37], and we assume this is due to the difficulty in the selection and delineation of the target area. The CT imaging of NTM pulmonary disease is complex and variable, including cavity signs, bronchiectasis, bronchitis, pulmonary consolidation, etc., the lesions exist alone or at the same time. These CT features were similar to pulmonary tuberculosis and difficult to distinguish. Some studies showed that the signs of bronchiectasis and cavity in CT images of NTM pulmonary diseases were different from those of pulmonary tuberculosis [28–30]. In this study, 62 of the 73 patients with NTM pulmonary disease were complicated with bronchiectasis, 58 of the 69 patients with NTM pulmonary disease were complicated with bronchiectasis. Since cavities were easy to observe and outline. besides, studies have shown that cavity has significant associations with mortality [38, 39], disease progression, and the failure of clinical treatment [40, 41],

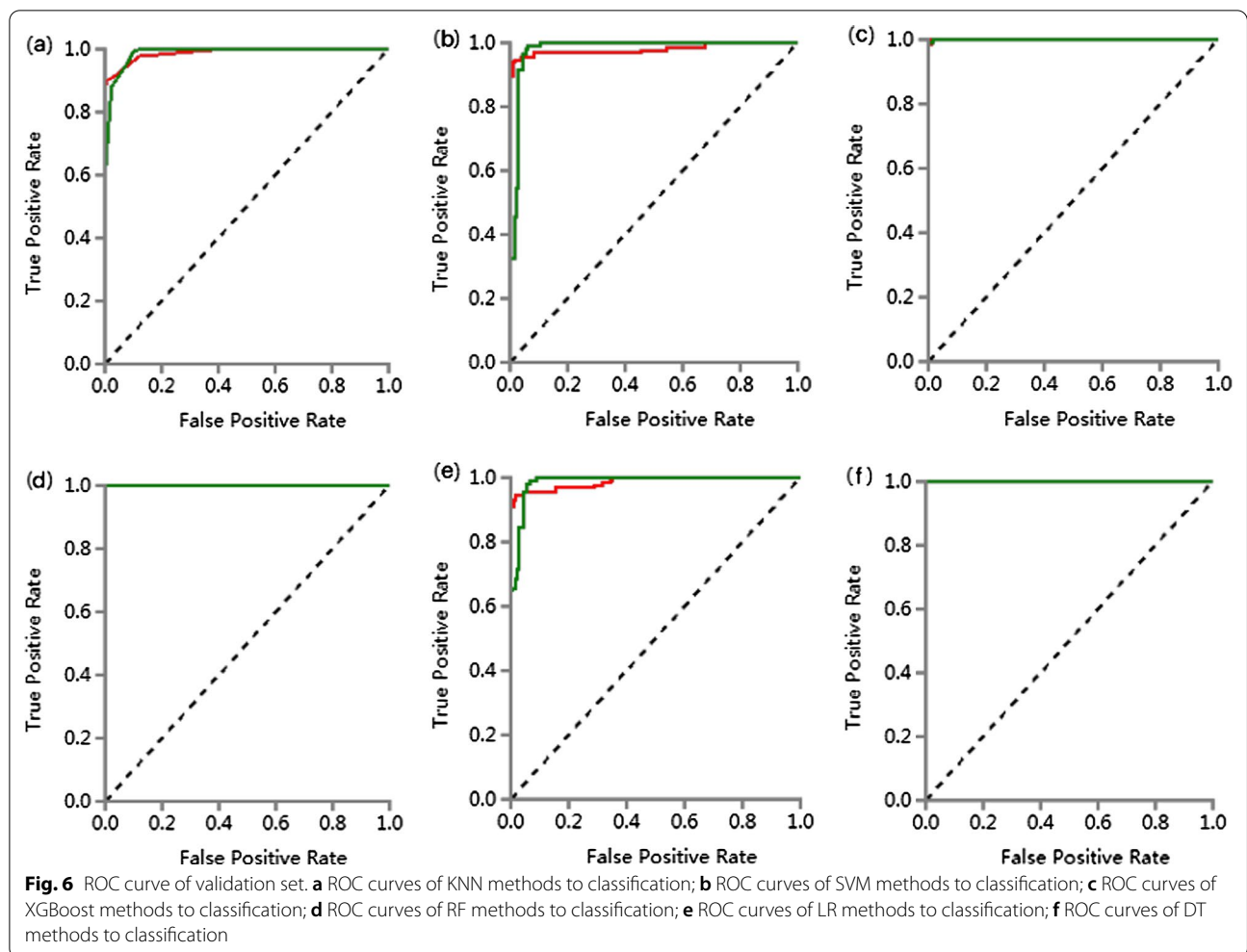


Table 2 ROC results with six classifiers of validation set

Classifiers	Category	AUC	95% CI	Sensitivity	Specificity
KNN	NTM	0.99	0.93–1.00	0.94	0.96
	TB	0.99	0.93–1.00	0.96	0.94
SVM	NTM	0.98	0.91–1.00	0.97	0.82
	TB	0.98	0.91–1.00	0.82	0.97
XGBoost	NTM	0.98	0.94–1.00	0.97	0.96
	TB	0.98	0.94–1.00	0.96	0.97
RF	NTM	0.99	0.93–1.00	0.94	0.96
	TB	0.99	0.93–1.00	0.96	0.94
LR	NTM	0.98	0.91–1.00	0.94	0.93
	TB	0.98	0.91–1.00	0.93	0.94
DT	NTM	0.85	0.76–0.94	0.91	0.79
	TB	0.85	0.76–0.94	0.79	0.91

hence, the indication by the international guidelines for the diagnosis and treatment of NTM pulmonary disease indicate that patients with cavity need active treatment [42]. Therefore, in this study, cavities were used as

the target areas for radiomics analysis, and the radiomics features obtained from cavities were used to explore the significance of identifying the disease from PTB. Radiomics can extract valuable cavity radiomics information from CT images for the differential diagnosis of the two diseases, which can improve the early diagnostic rate of NTM pulmonary disease and prevent misdiagnosis and mistreatment. The Xpert and PCR laboratory tests are DNA tests of *Mycobacterium tuberculosis* used to diagnose tuberculosis infection, but cannot diagnose NTM infection. Radiomics can be used in conjunction with the laboratory tests such as Xpert and PCR to prevent early and efficient diagnosis and treatment. Early diagnosis of NTM pulmonary disease can improve the quality of the patient’s life and promote the treatment of diseases, especially for the resource deficient health care system in developing countries [43, 44].

In this study, the cavities of NTM pulmonary disease and PTB were used as potential areas of interest for delineation. A total of 29 radiomics features were obtained from ROIs, of which 14 were texture analysis

Table 3 The results of four indicators—precision, recall, F1-score, support in training set

	Indicators	KNN	SVM	XGBoost	RF	LR	DT
TB	Precision	0.90	0.95	0.98	0.99	0.94	1.00
	Recall	0.96	0.96	0.99	1.00	0.95	1.00
	F1-score	0.93	0.95	0.99	1.00	0.95	1.00
	Support	108	108	108	108	108	108
NTM	Precision	0.97	0.97	0.99	1.00	0.96	1.00
	Recall	0.92	0.95	0.98	0.99	0.95	1.00
	F1-score	0.94	0.96	0.99	1.00	0.96	1.00
	Support	131	131	131	131	131	131

Table 4 The results of four indicators—precision, recall, F1-score, support in validation set

	Indicators	KNN	SVM	XGBoost	RF	LR	DT
TB	Precision	0.93	0.96	0.96	0.93	0.93	0.88
	Recall	0.96	0.82	0.96	0.96	0.93	0.79
	F1-score	0.95	0.88	0.96	0.95	0.93	0.83
	Support	28	28	28	28	28	28
NTM	Precision	0.97	0.86	0.97	0.97	0.94	0.83
	Recall	0.94	0.97	0.97	0.94	0.94	0.91
	F1-score	0.95	0.91	0.97	0.95	0.94	0.87
	Support	33	33	33	33	33	33

Table 5 ROC results with six classifiers of external validation set

Classifiers	Category	AUC	95% CI	Sensitivity	Specificity
KNN	NTM	0.84	0.66–0.82	0.65	0.83
	TB	0.84	0.66–0.82	0.83	0.65
SVM	NTM	0.90	0.78–0.92	0.94	0.77
	TB	0.90	0.78–0.92	0.77	0.94
XGBoost	NTM	0.96	0.75–0.87	0.97	0.65
	TB	0.96	0.75–0.87	0.65	0.97
RF	NTM	0.84	0.64–0.80	0.61	0.83
	TB	0.84	0.64–0.80	0.83	0.61
LR	NTM	0.95	0.84–0.96	0.94	0.87
	TB	0.95	0.84–0.96	0.87	0.94
DT	NTM	0.98	0.78–0.89	0.71	0.97
	TB	0.98	0.78–0.89	0.97	0.71

and 15 were first-order statistical feature groups. The study used six supervised learning classifiers (KNN, SVM, XGBoost, RF, and LR and DT) to analyze the extracted lung cavity features. In the training set, the AUC values of the six models were all greater than 0.98, 95% CI were 0.95–1.00, the sensitivity was greater than 0.92, and the specificity was greater than 0.92. In the validation set, the AUC values of the six models were

all greater than 0.85, 95% CI were 0.76–1.00, the sensitivity was greater than 0.79, and the specificity was greater than 0.79. The obtained AUC values of ROC curve were all high, and their sensitivity and specificity were greater than 0.79. In the external validation set, The AUC value was similar to the results of internal validation, but the sensitivity, specificity and 95% CI were lower than those of internal validation. Among the six classifiers, LR classifier has highest sensitivity and specificity, which were 0.94, 0.87. The study analyzed the characteristics of the six classifiers through four clinical indicators (accuracy, recall, f1 points, and support). In the training set, the precision of the six models was greater than 0.90, the recall rate was greater than 0.92, the F1 score was greater than 0.93, and the support were 108 and 131. In the validation set, the precision of the six models was greater than 0.83, the recall rate was greater than 0.79, the F1 score was greater than 0.83, and the support were 28 and 33. In the external validation set, the precision of the six models was greater than 0.57, the recall rate was greater than 0.61, the F1 score was greater than 0.68. Furthermore, the LR classifier has the highest precision, recall, and F1-score, which were 0.92, 0.94, 0.93 respectively in NTM group. The results showed that the radiomics features derived from the cavities could help in differentiating between

Table 6 The results of four indicators—precision, recall, F1-score, support in external validation set

	Indicators	KNN	SVM	XGBoost	RF	LR	DT
TB	Precision	0.60	0.88	0.63	0.57	0.90	0.67
	Recall	0.83	0.77	0.97	0.83	0.87	0.97
	F1-score	0.69	0.82	0.76	0.68	0.88	0.79
	Support	30	30	30	30	30	30
NTM	Precision	0.86	0.87	0.97	0.86	0.92	0.97
	Recall	0.65	0.94	0.65	0.61	0.94	0.71
	F1-score	0.74	0.90	0.78	0.71	0.93	0.82
	Support	49	49	49	49	49	49

the NTM pulmonary disease and PTB. Although some studies have reported that the CT imaging characteristics of NTM pulmonary disease cavities observed through traditional clinical methods differ from those of PTB [17, 45], the results obtained by researchers' naked eye recognition may appear subjective. Consequently, the effect is significantly reduced in actual clinical applications. The radiomics analysis of the cavity characteristics of pulmonary disease has the advantages of objectiveness, quantification, stability, and nonexperience dependence. Therefore, it presents a significant value for clinical application. It can be seen that the use of the radiomics characteristics in differentiating NTM pulmonary disease from PTB is very promising, and the results may help clinical radiologists lacking experience or have difficulty in visual recognition. Therefore, this technology is a very important, non-invasive, simple, and easy method in the identification of these two diseases.

In addition, although the pathological features of NTM pulmonary disease and tuberculosis are similar, there are some differences as follows; (1) the lesions of the NTM pulmonary disease have less caseous necrosis than PTB; (2) epithelial-like cells aggregation tends to be proliferative granuloma; and (3) the cavities are more common and the characteristics may be in multiple forms, multiple rooms and uneven inner walls [15]. Thus, the subtle differences in these pathological features of conventional CT images are difficult to be distinguished, resulting in a low level of differential diagnosis from PTB. The high-throughput extracted CT data of radiomics can reflect the internal differences of the pathological characteristics of the two diseases, which is consistent with the pathological performance and makes up for the lack of visual observation and key information in CT images.

However, this study also has some limitations: 1. to ensure the homogeneity of the images, images with 5 mm layer thickness were used in the research process.

Compared with thinner slice images, some information may be lost. 2. ROIs were delineated manually and individual subjectivity is unavoidable. 3. We focused only on the cavity as the research target and ignored other image features, which could lead to incomplete information and the inability to study some typical patients. 4. Increase the amount of data and further improve the data analysis model. The above shortcomings and deficiencies will be the focus of future research.

Conclusively, this study shows that the radiomics features based on CT imaging are effective in identifying NTM pulmonary disease cavities and PTB cavities, which is highly significant for early non-invasive diagnosis and differentiation of NTM pulmonary disease with cavity. Among the six classifiers, the LR classifier has the best performance in identifying the two diseases.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12890-021-01766-2>.

Additional file 1. Data model and results in this study.

Additional file 2. External validation of the model in this study.

Additional file 3. Ethical proof.

Acknowledgements

The authors gratefully acknowledge the financial supports by the Foundation of Health Commission of Shandong Province; Shandong medical and Health science and technology development plan project number: 2019WS535.

Authors' contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Funding

Organization: Health Commission of Shandong Province; Entry name: Shandong medical and health science and technology development plan project; Project number: 2019WS535; Project Leader: Qinghu Yan.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

All authors attest that they meet the Helsinki declaration. The authors declare that they consent to participate.

Consent for publication

Not applicable.

Informed consent

This retrospective studies included all subjects are under 18 and informed consent obtained from a parent and/or legal guardian.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Radiology, Shandong Public Health Clinical Center, Jinan 250013, China. ²Department of Radiology, Qilu Hospital of Shandong University, Jinan 250012, China. ³Huiying Medical Technology (Beijing) Co., Ltd, Beijing 100192, China.

Received: 27 April 2021 Accepted: 24 November 2021

Published online: 07 January 2022

References

- Kendall BA, Winthrop KL. Update on the epidemiology of pulmonary nontuberculous mycobacterial infections. *Semin Respir Crit Care Med*. 2013;34(1):87–94. <https://doi.org/10.3201/eid2206.151086>.
- Simons S, van Ingen J, Hsueh PR, et al. Nontuberculous mycobacteria in respiratory tract infections, eastern Asia. *Emerg Infect Dis*. 2011;17(3):343–9. <https://doi.org/10.3201/eid1703.100604>.
- Chinese Medical Association Tuberculosis Branch, Editorial Board of Chinese Journal of Tuberculosis and Respiratory Diseases. Expert consensus on the diagnosis and treatment of nontuberculous mycobacterial disease. *Chin J Tubercul Respir Dis*. 2012;35(8):572–580. <https://doi.org/10.3760/cma.j.issn.1001-0939.2012.08.006>
- Peiming C, Fengfeng L, Yan L, et al. Clinical analysis of 58 cases of nontuberculous mycobacteria. *Chongqing Medical*. 2014;43(7):854–6. <https://doi.org/10.3969/j.issn.1671-8348.2014.07.032>.
- Gopalaswamy R, Shanmugam S, Mondal R, Subbian S. Of tuberculosis and non-tuberculous mycobacterial infections—a comparative analysis of epidemiology, diagnosis and treatment. *J Biomed Sci*. 2020;27(1):74.
- Chu HQ, Li B, Zhao L, et al. Chest imaging comparison between nontuberculous and tuberculosis mycobacteria in sputum acid fast bacilli smear-positive patients. *Eur Rev Med Pharmacol Sci*. 2015;19(13):2429–39.
- Li F, He W, Zhou X, et al. The similarities and differences of HRCT imaging features between nontuberculous mycobacterial lung diseases and active pulmonary tuberculosis. *J Chin Anti-Tuberc Assoc*. 2018;40(5):499–505. <https://doi.org/10.3969/j.issn.1000-6621.2018.05.012>.
- Jingbo Z, Haiqing C. Clinical and imaging features of nontuberculous mycobacterial lung diseases. *Chin J Infect Chemother*. 2016;16(1):86–91. <https://doi.org/10.16718/j.1009-7708.2016.01.020>.
- Yuan MK, Chang CY, Tsai PH, Lee YM, Huang JW, Chang SC. Comparative chest computed tomography findings of non-tuberculous mycobacterial lung diseases and pulmonary tuberculosis in patients with acid fast bacilli smear-positive sputum. *BMC Pulmonary Med*. 2014;14(1).
- Lange C, Mori T. Advances in the diagnosis of tuberculosis. *Respirology*. 2010;15:220–40.
- Kim C, Park SH, Oh SY, et al. Comparison of chest CT findings in nontuberculous mycobacterial diseases vs. Mycobacterium tuberculosis lung disease in HIV-negative patients with cavities. *PLoS ONE*. 2017;12(3):e0174240.
- Hui G. To explore the clinical value of CT in the diagnosis and clinicopathological classification of peripheral small lung cancer. *China Continuing Med Educ*. 2016;8(21):48. <https://doi.org/10.3969/j.issn.1674-9308.2016.21.027>.
- Ma J, Zhou Z, Ren Y, et al. Computerized detection of lung nodules through radiomics. *Med Phys*. 2017;44(8):4148–58.
- Coroller TP, Agrawal V, Huynh E, et al. Radiomic-based pathological response prediction from primary tumors and lymph nodes in NSCLC. *J Thorac Oncol*. 2017;12(3):467–76.
- Wittram C, Weisbrod GL. Mycobacterium avium complex lung disease in immunocompetent patients: radiography-CT correlation. *Br J Radiol*. 2002;75(829):340–4.
- Chinese Medical Association Tuberculosis Branch. Guidelines for the diagnosis and management of nontuberculous mycobacteria. *Chin J Tubercul Respir Dis*. 2000;23(11):650–3. <https://doi.org/10.3760/j.issn.1001-0939.2000.11.003>.
- Daley CL, Iaccarino JM, Lange C, et al. Treatment of nontuberculous mycobacterial pulmonary disease: an official ATS/ERS/ESCMID/IDSA clinical practice guideline. *Eur Respir J*. 2020;56(1):200053. <https://doi.org/10.1183/13993003.00535-2020>.
- Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441–6. <https://doi.org/10.1016/j.ejca.2011.11.036>.
- Wang L, Ding W, Mo Y, Shi D, Zhang S, Zhong L, Wang K, Wang J, Huang C, Zhang S, Ye Z, Shen J, Xing Z. Distinguishing nontuberculous mycobacteria from Mycobacterium tuberculosis lung disease from CT images using a deep learning framework. *Eur J Nucl Med Mol Imaging*. 2021. <https://doi.org/10.1007/s00259-021-05432-x>.
- Falkinham JO 3rd. Nontuberculous mycobacteria in the environment. *Clin Chest Med*. 2002;23(3):529–51.
- Aboagye SY, Danso E, Aboagye SY, et al. Isolation of nontuberculous mycobacteria from the environment of Ghanaian communities where buruli ulcer is endemic. *Appl Environ Microbiol*. 2016;82(14):4320–9.
- Chen CY, Chen HY, Chou CH, et al. Pulmonary infection caused by nontuberculous mycobacteria in a medical center in Taiwan, 2005–2008. *Diagn Microbiol Infect Dis*. 2012;72(1):47–51.
- Vinnard C, Longworth S, Mezocho A, et al. Deaths related to nontuberculous mycobacterial infection in the United States, 1999–2014. *Ann Am Thorac Soc*. 2016;13(1):1951–5.
- Ringshausen FC, Wagner D, De Roux A, et al. Prevalence of nontuberculous mycobacterial pulmonary disease, Germany, 2009–2014. *Emerg Infect Dis*. 2016;22(6):1102–5.
- Marras TK, Mendelson D, Marchand AT, et al. Pulmonary nontuberculous mycobacterial disease, Ontario, Canada, 1998–2010. *Emerg Infect Dis*. 2013;19(11):1889–91.
- Kwak N, Lee CH, Lee HJ, Kang YA, Lee JH, Han SK, Yim JJ. Non-tuberculous mycobacterial lung disease: diagnosis based on computed tomography of the chest. *Eur Radiol*. 2016;26(12):4449–56. <https://doi.org/10.1007/s00330-016-4286-6>.
- Aksamit TR, Phillely JV, Griffith DE. Nontuberculous mycobacterial (NTM) lung disease: the top ten essentials. *Respir Med*. 2014;108(3):417–25.
- Eisenberg I, Yasin A, Fuks L, Stein N, Saliba W, Kramer MR, Adir Y, Shteinberg M. Radiologic characteristics of non-tuberculous mycobacteria infection in patients with bronchiectasis. *Lung*. 2020;198(4):715–22. <https://doi.org/10.1007/s00408-020-00371-0>.
- Kwak N, Lee JH, Kim HJ, Kim SA, Yim JJ. New-onset nontuberculous mycobacterial pulmonary disease in bronchiectasis: tracking the clinical and radiographic changes. *BMC Pulm Med*. 2020;20(1):293. <https://doi.org/10.1186/s12890-020-01331-3>.
- Kim C, Park SH, Oh SY, Kim SS, Jo KW, Shim TS, Kim MY. Comparison of chest CT findings in nontuberculous mycobacterial diseases vs. Mycobacterium tuberculosis lung disease in HIV-negative patients with cavities. *PLoS ONE*. 2017;12(3):e0174240.
- Jie D, Jingyun S, Li L, et al. Imaging findings of nontuberculous mycobacterial lung diseases: a comparison with findings of secondary pulmonary tuberculosis. *J Chin Antituberc Assoc*. 2014;36(8):706–9.
- Liu Y, Kim J, Balagurunathan Y, et al. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clin Lung Cancer*. 2016;17(5):441–448.e6. <https://doi.org/10.1016/j.clcc.2016.02.001>.
- Ganesan B, Panayiotou E, Burnand K, et al. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur Radiol*. 2012;22(4):796–802. <https://doi.org/10.1007/s00330-011-2319-8>.

34. Gevaert O, Xu J, Hoang CD, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. *Radiology*. 2012;264(2):387–96. <https://doi.org/10.1148/radiol.12111607>.
35. Zhang T, Yuan M, Zhong Y, et al. Differentiation of focal organising pneumonia and peripheral adenocarcinoma in solid lung lesions using thin-section CT-based radiomics. *Clin Radiol*. 2019;74(1):78.e23–78.e30. <https://doi.org/10.1016/j.crad.2018.08.014>.
36. Yanling W, Duo G, Zuojun G, et al. Radiomics nomogram analyses for differentiating pneumonia and acute paraquat lung injury. *Sci Rep*. 2019;9(1):15029. <https://doi.org/10.1038/s41598-019-50886-7>.
37. Yanling W, Duo G, Zuojun G, et al. Radiomics nomogram analyses for differentiating pneumonia and acute paraquat lung injury. *Sci Rep*. 2019;9(1):15029.
38. Hayashi M, Takayanagi N, Kanauchi T, et al. Prognostic factors of 634 HIV-negative patients with Mycobacterium avium complex lung disease. *Am J Respir Crit Care Med*. 2012;185(5):575–83. <https://doi.org/10.1164/rccm.201107-1203OC>.
39. Gochi M, Takayanagi N, Kanauchi T, et al. Retrospective study of the predictors of mortality and radiographic deterioration in 782 patients with nodular/bronchiectatic Mycobacterium avium complex lung disease. *BMJ Open*. 2015;5: e008058. <https://doi.org/10.1136/bmjopen-2015-008058>.
40. Rosenzweig DY. Pulmonary mycobacterial infections due to Mycobacterium intracellulare-avium complex. Clinical features and course in 100 consecutive cases. *Chest*. 1979;75:115–9. <https://doi.org/10.1378/chest.75.2.115>.
41. Lam PK, Griffith DE, Aksamit TR, et al. Factors related to response to intermittent treatment of Mycobacterium avium complex lung disease. *Am J Respir Crit Care Med*. 2006;173:1283–9. <https://doi.org/10.1164/rccm.200509-1531OC>.
42. Cowman SA, Jacob J, et al. Latent class analysis to define radiological subgroups in pulmonary nontuberculous mycobacterial disease. *BMC Pulm Med*. 2018;18(1):145. <https://doi.org/10.1186/s12890-018-0675-8>.
43. Tan Y, Su B, Shu W, et al. Epidemiology of pulmonary disease due to nontuberculous mycobacteria in Southern China, 2013–2016. *BMC Pulm Med*. 2018;18(1):1–7.
44. Maurya AK, Nag VL, Kant S, et al. Prevalence of nontuberculous mycobacteria among extrapulmonary tuberculosis cases in tertiary care centers in Northern India. *Biomed Res Int*. 2015;2015:5.
45. Lu C, Heshui S, Benmei L, et al. Comparative analysis of multi-slice spiral CT signs of nontuberculous mycobacterial lung disease and secondary tuberculosis. *J Clin Radiol*. 2019;38(8):1400–4.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

